EOSC-SYNERGY

EU DELIVERABLE: D2.1

Roadmap for integration of national capacities into the EOSC

and Policy Gap analysis

Document Identifier:	EOSC-SYNERGY-D2.1
Date:	29/02/2020
Activity:	WP2
Leader Partner:	AGH/AGH-UST
Document Status:	APPROVED
Dissemination Level:	PUBLIC
Document Link:	https://drive.google.com/open?id=1OEDrZIN6YnE_yoT5Pt25IVYVBL-Hrv4v

Abstract:

This deliverable summarises the considerations made and decisions taken to arrive at the initial roadmap for the integration of national capacities.

To provide thematic services with available infrastructure as early as possible, the majority of the resources will be integrated with high priority with the EGI Federated Cloud, while the more work intensive components will be integrated after the “low hanging fruit” have already been integrated.

The current status of adoption of the AARC Policy Pack across the infrastructure was assessed with a questionnaire, results of which are provided. Room for improvement was identified and will be addressed with targeted online training on identified gaps.

As a result of this deliverable, the majority of the infrastructure can be integrated with EOSC via the EGI Federated Cloud, while a smaller part of the capacities requires a case by case handling and adaptation / integration of software solutions into the operation of the participating infrastructure partners.

I. Copyright Notice

II. Delivery Slip

	Name	Partner/Activity	Date
From	Marcus Hardt Agnieszka Pułapa	KIT CYFRONET	08/03/2020
Reviewed by	Moderator: Isabel Campos Reviewers: Marcin Plociennik Alberto Azevedo	CSIC PSNC LNEC	11/03/2020
Approved by	PMB	ALL	23/03/2020

III. Document Log

Issue	Date	Comment	Author/Partner
v 1.0	31.01.2020	First draft	A.Pułapa/CYFRONET
v 2.0		EOSC-hub service integration and Infrastructure services-integration status	M. Hardt/KIT, A.Vieira/LIP, J.Pina/LIP, I.Diaz.CESGA, A.Coveney/STFC M.Hardt/KIT, A.Lopez/CSIC, M.David/LIP, M.Orzechowski/CYFRONET, M.Pospieszny/PSNC, V.Tran/IISAS ,I.Krenkova/CESNET, R.Diez/CESGA,D.Martinez/BIFI ,A.Pardo /CETA-CIEMAT
v 3.0	26.02.2020	Policy Gap Analysis	U. Stevanovic/KIT
v 4.0	28.02.2020	First Draft for internal review	M.Hardt/KIT, A.Pułapa/CYFRONET
v 5.0	10.03.2020	Final Document	M.Hardt/KIT, A.Pułapa/CYFRONET

IV. List of Acronyms

The full list of acronyms is in Appendix A.

Table of Contents

1. Executive Summary

This report is the first deliverable of the Infrastructure Work Package of EOSC-Synergy. Its goal is to shape the future functioning of the computing and storage resources provided by participating national and thematic capacities. The main focus lies on the integration with similar infrastructures operated in the context of the EOSC. It describes the guiding principles that were identified so that the integration with EOSC can be steered appropriately. It was found that the integration is not straightforward at the moment, because different levels of integration are possible, yet advertised differently. WP2 represents the national and thematic capacities and also the repositories. We found that their integration needs to be done on two different levels: While the higher level services, i.e. the repositories, need to be integrated via the so called “onboarding procedure” of EOSC, the cloud services that constitute the capacities need to be integrated with the EGI Federated Cloud, and remaining HPC services need to be integrated via individual integration paths.

In addition, integration on the policy level is required. The technical policy integration (security, trust, ...) is part of this report, while the organisational and international policy questions are part of the work in WP5.

To ramp-up the project’s infrastructure as fast as possible, we prioritised the infrastructure integration over the onboarding processes. Most of the individual capacities are already integrated with the EGI Federated Cloud. The report details the depth of integration that is done at the time of writing. In addition, sites report the capacities provided and the Virtual Organisations supported.

In parallel, the policy group reports about the questionnaire circulated across the computer centres, to evaluate the degree of support of the technical policies in infrastructures that the AARC/AARC2 projects have generated in their Policy Development KIT. We present the result of the questionnaire and identify the need for training so the awareness of and compliance to the policies can be increased.

1. Introduction

EOSC-Synergy is a collaborative project involving several institutions from different European countries working together to combine their knowledge and expertise to expand the capability and capacity of EOSC.

This report identifies the roadmap for the integration of national capacities into EOSC. This roadmap is broken down into the individual national integration plans and includes the amounts of resources provided per national service.

The guiding principles we adhered to are

- Sustainability: We are fully aware that a project like EOSC-Synergy with a runtime of 30 months, can not build a sustainable infrastructure. Our efforts were therefore bundled on
    -

- Understanding the current offerings of EOSC, which are very diverse, which made it difficult to find the appropriate offering for the tasks of EOSC-Synergy.
- Identifying the steps and the level of integration that are required to adequately support the demanding use-cases of EOSC-Synergy
- Creating a tangible set of documentation that is easy to follow for the participating national capacities, so they can be integrated quickly.

- Seamlessness: To be useful for the multitude of participating thematic services and repositories, the integration had to address the ease-of-use of the resources at the same time as we could not change their existing computing models. A major boundary condition for the design of the integration roadmap therefore involved a close collaboration with the use-cases of WP4.
- Expansion by integration: Similarly, the national and thematic resource providers are participating primarily in national and thematic computing projects. As such, they are unable to redesign their offerings to fulfil particular demands of EOSC-Synergy. This is why the roadmap took an integrative approach, that extends existing infrastructures, without changing the way in which they are currently delivering their services.
- Wholistic Policy support: Many existing services operated in their own nation or in their own domain. When integrating such services into a federated environment such as EOSC, several organisational and legal topics (i.e. policies) have to be addressed. Close collaboration with WP5 helps us to include the existing service providers into the federated (policy) world.

Based on these principles, the level of integration with EOSC had to be determined. We came to the conclusion that the so-called “onboarding process” of EOSC Hub, which integrates services into a manually controlled environment called marketplace, was targeted to services such as repositories and thematic services. However, this was not deemed appropriate for the integration of national and thematic capacities. This is because for a useful integration, a deeper technical with automated procedures are required. Therefore, and in order to have the largest possible impact within the time given, we had to prioritise the services to be integrated and on which level the integration had to be done.

To use the “low hanging fruits” first, we decided to first integrate the existing cloud infrastructures with the EGI Federated Cloud [EGI-Fedcloud], because Federated Cloud is well understood and because most of the participating sites of EOSC-Synergy already had the required OpenStack (or OpenNebula) services running. The large benefit of the Federated Cloud, as operated by EGI, is that much operational support is provided by an institution with a longer-term mission. In addition, the experience in operation the Federated Cloud generates several advantages for the participating national and thematic capacities. For example, the provided external monitoring will help to improve the services provided by the sites. The also provided accounting can help in generating a better understanding of how resources are allocated to the different user groups at a centre. The well-established security procedures will support a better reaction to security incidents.

Based on the experience with the integration of standard services into a well understood platform, in the future, additional services (such as HPC computing and storage resources), will be integrated. Their integration is more demanding in terms of time and software components required, because we currently

see a transformation from a classical distributed computing model towards cloud models. This is important because classic HPC centres are much more eager for inclusion into a federated environment under the cloud paradigm. However, of-the-shelf solutions, such as OpenStack, do not exist and can, in many cases, not cater for the complexity and heterogeneity of HPC systems.

Policy integration plays an important aspect for the integration of formerly isolated regional, national or thematic capacities into a global endeavor such as EOSC. The initial scope within the reporting period is to understand the current adoption of policies for federated operation. The baseline for this is the policy development kit (PDK), provided to us by the AARC/AARC2 EU-Projects. This lays the ground for raising the awareness of the policies, but also to identify the gap between the existing AARC PDK [AARC PDK] and the reality in today's capacities. The gap analysis will identify focal points for future training events as part of EOSC-Synergy.

1.1 Scope of the document

This report analyses the status of the national and thematic capacities to be integrated to identify gaps that need to be overcome for full interoperation with existing services in the EOSC context. The EOSC-Synergy services are being analysed in terms of their architectural type and as a result of this analysis, the possible ways of integration with the EOSC core services will be considered. This document elaborates on the harmonisation activities required to achieve this initial integration.

1.2 Document organisation

This document is comprised of 7 sections. After this introduction, Section 2 identifies core services by their functions to integrate and provide the initial procedures for the integration. Section 3 provides an overview of the infrastructure services of the project, covering a description of the service, current status and further plans for integration. Section 4 identifies Policy Gap Analysis and section 5 includes information about the Gender Issues. Finally, section 7 covers the generic document templates. Appendix A contains a full list of acronyms used in the document. Appendix B includes the table with a summary of service integration status. Appendices C and D will be focused on the Policy Questionnaire.

2. EOSC-hub services integration

The overall goal for the integration is to support building a sustainable infrastructure. Therefore, all national capacities will be guided to support their infrastructures in a way that it does not depend on EOSC-Synergy, but on a federated EOSC-enabled solution following the guidelines being developed in the EOSC initiative instead. This is the required condition for a sustainable approach so that national capacities will remain available after the end of EOSC-Synergy.

The general strategy for integration with EOSC is driven on multiple levels.

Level 1: From the perspective of the cloud related infrastructure the first step is to integrate with the EGI Federated Cloud [EGI-FedCloud, EGI-FedCloud-Architecture] which is a central component of EOSC.

EOSC-Synergy will integrate all resources supported with the Federated Cloud, which covers mostly Computing (i.e. Virtual Machines (e.g. OpenNebula, OpenStack)) and Storage, such as OpenStack Swift.

Level 2: Integration of those services that require a manual extension for integration with core services. This may include for example specific storage or computing resources that cannot easily be integrated via Federated Cloud. Examples inlcude HPC clusters or large storage installations with limitations on the services that can be operated / supported by the national sites. In particular, the repositories are complex services that require deep integration both with underlying (AAI, Monitoring, ...) servies, but also with higher level services.

Level 3: EOSC-Synergy will make use of these infrastructures to provide higher level services, such as the repositories and the thematic services of WP4. The integration on this level targets the customisation to user requirements and is generally guided by the EOSC Onboard Procedure [EOSC Onboarding]. This also involves leveraging the EOSC Marketplace to attract larger amounts of users to the services provided there.

Our work starts with the integration on the infrastructure level with the Federated Cloud. The progress of the individual sites is outlined in the respective subsection of the sites.

2.1 Identified services by their functions to integrate

In cooperation with EOSC we've already identified the initial list of federated services and their functions and we prepared a document as a point of reference [EOSC core and common]. EOSC-Synergy services were analyzed in terms of their architectural service type and service delivery level to recognize possible integrations with EOSC services. Collected requirements were used to select the first federated services to integrate and identify common integration scenarios. After thorough analysis, five fundamental federated services were chosen for the first implementation phase, as follows: AAI, Monitoring, Accounting, Information Provider and Service Registry.

2.2 Initial procedures for integration

In this section, a description of the benefits associated with the integration of the five chosen federated services and procedures created to help the integration of EOSC-Synergy services in EOSC, is presented. Concrete information about integration status is described in section 3.

2.2.1 *AAI*

The integration with the Authentication and Authorisation Infrastructure (AAI) is the fundamental step, because this is the mechanism that delivers the identities (unique identifiers and group memberships) to the infrastructure. This is essential in order for the users to access the services in the first place. Furthermore, AAI is the logical "point" where users may be informed about the policies (like privacy notices, AUPs) and where access rights to services may be enforced.

In collaboration with EOSC-Hub and EGI we have recorded a procedure [AAI-Integration-Procedure]. There, services are first integrated with a demonstration instance of the EGI Check-in service for initial

testing. Successfully tested services will then be moved to the production instance of EGI Check-in. This process is documented and supported in various procedures of EGI [EGI-Proc-09].

This integration will allow users to access all services, using their home-Identity, their community identities supported by EGI (six at the time of writing) or one of six social identity providers. In addition, Virtual Organisations (VOs) have been created [EGI-Proc-14] to support the thematic services and facilitate their resource allocation requests. For conducting trainings we envisage collaboration with the EGI training VO.

2.2.2 *Monitoring*

Every production service requires a monitoring system, so the desired efficiency and accountability can be verified at any time. After defining what should be monitored, it is easy to create a test and display its results. For EOSC-Synergy, Nagios core service will be used. One of the key aspects of Nagios, is to get the information about a particular test.

With Nagios we can monitor various kinds of hosts, services and actions. From the simplest check on a web service (ping) to a more complex test on the service itself, such as testing outputs giving some input on APIs, websites using selenium or even test if the login runs smoothly. This is all examples, since plugins can be written to all types of tests to check even the small functionality, using different programming languages.

In addition, contacts and alerts will be used to notify about problems detected by the Nagios service. This allows us to efficiently react to problems in a timely manner.

2.2.3 *Accounting*

The EOSC Accounting service collects, stores, aggregates, and displays usage information of HTC compute, storage space, cloud VM and data set resources. Resource Centres that are providing compute or storage to the EOSC infrastructure have to implement a collector (a stand-alone script or program, or a built-in function of their resource system), that gathers accounting metrics formatted into a standardised record format. These metrics are then transferred via a messaging service to the Accounting Repository, which stores and processes the data to produce aggregations that are then sent to the Accounting Portal for display.

The Accounting Portal retrieves topology information on how resource centres relate to national infrastructures and regions from the configuration management database (CMDB) and community affiliation from the AAI service to properly organise the accounting data. Information related to groups or VOs should also contain information about scientific disciplines to allow the portal to properly classify the resource usage. The Accounting Portal already is integrated with several other tools, such as GOCDB and REBUS for topology and geographical data, Check-In for the AAI, the Operations Portal for VO and scientific discipline information. It also uses X.509 certificates to map users to institutions and these to countries.

2.2.4 *Information Provider*

The information system collects data from the resource providers in a research infrastructure and makes it available for workload orchestration. In the EGI e-Infrastructure this is a fundamental service both for the

HTC or Grid technology and for the EGI FedCloud. There are different solutions for each type of service within those technologies, e.g. both ARC and HTCondorCE have adhoc provider implementations for gathering the information. For Federated Cloud, we use a unique implementation, coined as cloud-info-provider, to fetch data from the supported Cloud Management Frameworks (CMFs), notably OpenStack and OpenNebula. The cloud-info-provider component leverages the APIs exposed by those CMFs to get key information about the Cloud resource provider, such as the projects and images that any given VO is allowed to use.

Consequently, the data collected by the information providers is essential for the operation of the different workload management services available through EOSC, such as the INDIGO PaaS Orchestrator or DIRAC4EGI.

2.2.5 *Service registry*

In EOSC the service registry is done in the Service Portfolio Management Tool (SPMT) which stores the service information required and previously validated by EOSC Service Validation Board (SVB). The SPMT stores the service descriptions of ALL EOSC services, including the HUB core services used for running the EOSC federation together with relevant information and configuration details about the service components. SPMT also links to service instance entries in EOSC-hub Configuration Management Database (CMDB) currently consisted by the GOCDB (provided by EGI) and DPMT (provided by EUDAT) and under the control of the EOSC-hub Service Management. The SPMT also allows exporting service descriptions to other tools and service catalogues, such as the one to be established by the eInfraCentral project and EOSC catalogue (https://www.eosc-hub.eu/catalogue).

3. National capacity integration status

This section describes the national capacities. A description of each service is provided. EOSC-Synergy supports 19 services at 10 sites. At the beginning of each chapter we describe some basic information about each site, which is followed by a general service description including the current integration status and future plans for integration. The table that summarises the work on the integration of the national capacities is shown in Appendix B.

3.1 KIT

KIT participates with its computer centre and the climate research group on Ozone Assessment (O3AS). As such, the resources provided by KIT will mostly be pledged for the O3AS VO. Additional resources will be made available for testing and development.

3.1.1 LSDF Storage Facility (Storage)

The KIT Large Scale Data Facility (LSDF) is a regional storage hardware project. In collaboration with the Heidelberg University, we provide a high performance Petabyte Scale storage (10PB disk + Tape Backup) funded by the regional and national science ministries. It supports a given set of science

disciplines (Climatology, photon science, structural biology, hydrodynamics, engineering, ...) and facilitates a high-performance integration with KIT HPC and Cloud facilities.

3.1.1.1 *Current status*

Integration is foreseen via the WebDAV protocol. Currently integration of the WebDAV module of Apache with the AAI is under way. The challenges lie in the technical integration of OIDC and WebDAV, i.e. the integration with global OIDC identities and local Unix UIDs.

3.1.1.2 *Further plans for integration*

Once Integrated we want to advertise the EOSC approach to KIT service providers, so allow a more general service delivery. We will try to support access to our existing cloud and HPC resources via EOSC interfaces.

Integration with Monitoring, Accounting and Information Systems is on the roadmap.

3.1.2 WATTS (Token Translation)

WaTTS allows using any legacy service with federated identities, such as eduGain or google.

For this, WaTTS accepts federated identities (via OpenID Connect) and uses a plugin scheme to generate credentials for your service. This allows you to provide services that do not normally support federated identities to federated users.

3.1.2.1 *Current status*

Integration with the AAI is done. Integration with monitoring and accounting are not deemed necessary at this point. Monitoring is done site-local at KIT, and accounting does not appear to make sense, as the actual resource usage is minimal.

3.1.2.2 *Further plans for integration*

We will investigate in supporting VOMS extensions in the certificates delivered.

3.1.3 FEUDAL (Account provisioning)

The FEderated User Deployment portAL (FEUDAL) is a system for the deployment of user accounts into the local user management systems of multiple (federated) resources. FEUDAL fills the gap between federated users and "legacy" services, that require the creation of user accounts in service (and often the placement of credentials such as ssh-keys) before a user may log in.

3.1.3.1 *Current status*

FEUDAL is integrated with the AAI. Integration with KIT Monitoring is underway. Accounting does not appear to make sense.

3.1.3.2 *Further plans for integration*

The next step is the integration of services across the distributed infrastructure that require the creation of accounts.

3.2 CSIC

CSIC participation is implemented via the Instituto de Física de Cantabria (IFCA). IFCA develops a research line in Advanced Computing and operates several computing and storage resources, including Cloud, HTC and HPC.

IFCA is offering part of its resources based on a cloud service model within the EGI coordination (Federated Cloud of EGI). In this context, over the last years, CSIC has assumed the leadership and coordination of some of the development tasks of the EGI Federated Cloud integration modules, critical software being used by all the resource providers that are part of the infrastructure.

3.2.1 *IFCA Scientific Cloud Infrastructure (OpenStack)*

The IFCA Scientific Cloud is based on OpenStack, providing Infrastructure as a Service (compute, storage, network) and Platform as a Service (resource orchestration, machine learning as a service, serverless) access models.

3.2.1.1 *Current status*

The service is integrated in the EGI.eu Federated Cloud, with all services referred in section 2.1 and 2.2 in place. As such, IFCA is listed in the GOCDB, monitored by ARGO, its accounting is being pushed to the EGI.eu accounting repository and its services are being exploited by several EU research communities. Regarding authentication, IFCA supports several OpenID Connect and OAuth2 identity providers,namely: EGI Check-in, DEEP-Hybrid-DataCloud IAM, eXtreme-DataCloud IAM and INDIGO-DataCloud IAM.

3.2.1.2 *Further plans for integration*

We plan to integrate and support higher level service models (PaaS, SaaS) under a hybrid approach.

3.3 LIP

LIP is one of the members of the Portuguese National Distributed Computing Infrastructure (INCD) a digital research infrastructure that delivers compute and data services to the Portuguese scientific and academic communities. The INCD infrastructure services include cloud IaaS, HPC and HTC as well as the storage and data management services that support the cloud and HTC/HPC clusters.

INCD has two sites, one in Lisbon (called INCD-NCG) and one in Braga (called INCD-RDA). INCD-NCG site is in operation for several years while the INCD-RDA was recently deployed and is not yet fully operational.

3.3.1 INCD NCG Cloud

The INCD-NCG cloud IaaS service is based on OpenStack (currently Stein version). OpenStack offers the usual IaaS services: computing, storage and networking. The storage service is provided by Ceph exposing Object storage S3 and Swift APIs.

3.3.1.1 *Current status*

The service is currently undergoing integration into the EGI Federated Cloud service, this is achieved by integrating with all services referred in section 2. It is integrated with the EGI Check-in federated AAI service, through the OpenID Connect and OAuth2 protocols. It is registered in the GOCDB, monitored through the EGI ARGO service, and publishing information through the cloud-info-provider service, the publishing of accounting records into APEL is undergoing.

3.3.1.2 *Further plans for integration*

The next steps are to integrate the cloud IaaS under higher level services; PaaS and SaaS, through coordinated deployments of services such as the Infrastructure Manager (IM).

3.3.2 INCD NCG Farm

The INCD-NCG farm is a “classical” cluster with a batch system and scheduler based on Son of Grid Engine (SoGE). It has a Lustre distributed file system as underlying storage solution. The cluster has the following types of resources, High Throughput Computing (HTC) for single node job execution, High Performance Computing (HPC) for parallel MPI jobs and nodes with Nvidia GPUs for CUDA applications.

3.3.2.1 *Current status*

The INCD-NCG farm has been integrated in the EGI HTC infrastructure for several years, via Grid middleware. It exposes the cluster through a Grid Compute Element and Storage Element, one of the data movement protocol is WebDAV, and the authentication and authorization is accomplished through X.509 certificates and Virtual Organization Management System (VOMS) as attribute provider.

3.3.2.2 *Further plans for integration*

The future plans include following and possibly deploying newer protocols for federated AAI such as OIDC and OAuth2 and taking advantage of WaTTS service for any needed token translation.

3.3.3 INCD RDA Cloud (OpenStack and Ceph)

The INCD-NCG cloud IaaS service will be based on OpenStack offering the same underlying IaaS service as the INCD NCG cloud infrastructure.

3.3.3.1 *Current status*

The infrastructure is not yet deployed.

3.3.3.2 *Further plans for integration*

The infrastructure will follow the same integration path as the INCD NCG cloud, and achieve integration into the EGI Federated cloud.

3.3.4 INCD RDA FARM (Slurm and Lustre All compute and storage nodes)

The INCD-RDA farm is a "classical" cluster with a batch system and scheduler based on Slurm. It has a Lustre distributed file system as underlying storage solution. The cluster is a High Performance Computing (HPC) for parallel MPI jobs and nodes, it is also available for HTC workloads.

3.3.4.1 *Current status*

The cluster is in operation but not yet integrated into the EGI HTC/HPC compute service.

3.3.4.2 *Further plans for integration*

The integration of the cluster in the HTC/HPC will be done through the INCD NCG farm. The existing HTC Compute Element will have a new compute resource that is this new cluster. In this way, a minimal amount of services are necessary to accomplish the integration in EOSC.

3.4 CYFRONET

Academic Computer Centre CYFRONET AGH (http://www.cyfronet.pl/en), together with the Department of Computer Science AGH, focuses on scalable distributed systems, cross-domain computations in loosely coupled environments, knowledge management and support for life sciences. The CYFRONET infrastructure is both compatible and interoperable with existing European and worldwide Grid frameworks.

3.4.1 CYFRONET Onedata

CYFRONET's 2 most powerful machines include Prometheus and Zeus clusters. Prometheus, ranked in the TOP 500 list (November 2019) at 241st position consists of more than 2,200 servers with InfiniBand network with 56 Gbit/s capacity, connecting more than 55,000 Intel Haswell cores. These are accompanied by 279 TB RAM in total, and by two storage systems of 10 PB in total and 180 GB/s bandwidth. Prometheus has also been equipped with 144 NVidia Tesla GPGPUs. CYFRONET develops Onedata (http://www.onedata.org) for distributed data management and unified access, initially designed for PLGrid now supporting usage by initiatives including INDIGO-DataCloud, EGI-Engage, eXtreme Data Cloud and Helix Nebula Science Cloud PCP.

3.4.1.1 *Current status*

An instance of Onedata service (https://datahub.egi.eu) is fully integrated with some EGI components, including AAI (EGI Check-in production) and monitoring (Nagios). The AAI integration includes VO support. The service has been part of the EGI catalog for several years.

3.4.1.2 *Further plans for integration*

CYFRONET will continue to develop Onedata. It will continue to provide the instance https://datahub.egi.eu, and will upgrade it to the newest versions. Also, as the projects progress we plan to explore possibilities to integrate with Info Provider and Service registry.

3.5 PSNC

Poznan Supercomputing and Networking Center (PSNC) affiliated to the Institute of Bioorganic Chemistry of the Polish Academy of Sciences is an internationally known node of the European Research Area in the field of IT infrastructure of science and an important R&D center in the field of information and communication technologies (ICT). As a development centre of e-Infrastructure, PSNC designed and built the Metropolitan Network POZMAN, High Performance Computing Center and the national broadband network PIONIER, maintained and still developed by PSNC.

3.5.1 Cloud PSNC IaaS

PSNC IAAS Cloud is based on OpenStack and contains about 100+ nodes with 1/10GbEth interconnect with several data storage backends.

PSNC IaaS is built on top of OpenStack (currently Train). The system is deployed in two data centers as two independent instances with high-speed back end connections in form of multiple 100Gbit links. In total both instances sum up to 150 physical compute nodes with different configurations (256-512 GB memory per node, CPUs both AMD and Intel). Storage is provided in the form of Ceph cluster ~1PB in total for virtual machines and additional storage. There is also limited high speed SSD storage available (~100TB capacity) managed by ScaleIO software available for VMs as block storage. On top of the IaaS infrastructure PSNC is also operating a PaaS instance in the form of OpenShift platform. Currently both web-interface for OpenStack and API access is limited to local network or via VPN for external users.

3.5.1.1 *Current status*

To provide OpenStack IaaS and EGI AAI integration PSNC OpenStack is currently under process of deployment and configuration of EGI Compute Cloud interfaces which should provide EGI AAI - OpenStack keystone integration, and support for Accounting and Monitoring.

3.5.1.2 *Further plans for integration*

Due to PSNC internal set of requirements and policies, PSNC is currently in the process of development of an "In-house User Database Management" solution based on Keycloak, which will provide support for external IDP/AAI connection.

3.5.2 PSNC HPC cluster

PSNC Eagle HPC cluster contains 1038 nodes with 30544 cores and 2GB+ RAM/core, which translate to 1.25Pflops peak performance. Cluster worker nodes are connected using two interconnect networks: 10/1 Gbit/s Ethernet and 56Gbit/s InfiniBand. Data storage connected to HPC Cluster contains two domains: 1.5 PB GPFS file system for $HOME directories and software storage and 3.5 PB LUSTRE file system for I/O intensive usage (ie. scratch space).

Eagle contains a set of login nodes, EGI/WLCG Grid interfaces and EGI/WLCG enabled storage (DPM/WebDAV, Xrootd and dCache, about 1 PB in disks, about 5 PB in tapes).

3.5.2.1 *Current status*

Currently, EAGLE cluster is used by several EGI VOs from WLCG/HEP and other communities. Current EOSC integration plan assumes providing support for EOSC VOs by proper configuration of endpoints (work in progress).

3.5.2.2 *Further plans for integration*

Mentioned earlier, PSNC "In-house User Database Management" solution will also provide access via login nodes to users with credentials/tokens from external IDP/AAI related sources.

3.6 IISAS

IISAS has joined the EGI Federated Cloud from the beginning of the federated cloud infrastructure.

3.6.1 OpenStack

IISAS has two cloud OpenStack sites that are integrated to EGI Federated Cloud: the generic CPU site IISAS-FedCloud and the accelerated Cloud with GPU IISAS-GPUCloud. Both are operated in production and serve different VOs in EOSC.

3.6.1.1 *Current status*

Both sites are fully integrated in EGI Federated Cloud with all components: AAI (EGI Check-in production), image management (AppDB), information service (BDII), monitoring (ARGO), accounting (APEL) and registry (GOCDB). Both sites provide services for generic VOs (fedcloud.egi, eu, ops, dteam), training VO (training.egi.eu) also several thematic VOs (vo.nextgeoss.eu, d4science.org, vo.lifewatch.eu, enmr.eu).

3.6.1.2 *Further plans for integration*

Support for the generic VO for EOSC-Synergy (synergy.eosc.eu) and selected thematic VOs in EOSC-Synergy is planned. An upgrade to the newer OpenStack version is planned in the near future.

3.7 CESNET

CESNET participates with MetaCentrum Cloud, a fully operational cloud center, in EGI FedCloud. MetaCentrum Cloud, managed by CESNET and CERIT-SC, is a part of the national e-infrastructure e-INFRA CZ, which consists of HTC, HPC clusters, Hadoop/Spark cluster, HPC cloud (OpenStack), Ceph long-term storage operated by CESNET, and supercomputing center IT4I.

Besides cloud installation, additional resources like HPC clusters and Ceph storage can be provided to selected thematic services.

3.7.1 Cloud Infrastructure

MetaCentrum Cloud is an IaaS service, based on the OpenStack (Stein version), provides computing services by exposing OpenStack API and storage services (based on Ceph, providing S3 and Swift API).

3.7.1.1 *Current status*

The center is fully operational, it serves as a production site in EGI FedCloud with all components: AAI (EGI Check-in production https://www.egi.eu/services/check-in/), image management (AppDB), information service (BDII), monitoring, accounting, helpdesk, and registry (GOCDB). The web interface can be used on URL https://cloud2.metacentrum.cz. The service is registered in the EOSC Marketplace.

The service is available for ~10 VOs with EGI SLA (life/envri/social-science, ESA, training).

User authentication is based on OpenID Connect and OAuth2 protocols. Besides EGI Check-in AAI, users can also use ELIXIR AAI and national e-INFRA CZ AAI to access cloud services.

3.7.1.2 *Further plans for integration*

CESNET have several actions planned for the near future, such as: support for EOSC-Life AAI, PaNOSC AAI, Indigo/DataCloud AAI; support for PaaS services, preferably through the coordinated deployment of services like the Infrastructure Manager; the dedicated installation of IaaS, designed for processing sensitive/medical data; and the support for the generic VO for EOSC-Synergy (synergy.eosc.eu) and selected thematic VOs, we are involved in, in EOSC-Synergy.

3.8 CESGA

The main mission of “Centro de Supercomputación de Galicia” (CESGA) is to contribute to the advance of Science and Technology, through research and application of high-performance computing and

communications, as well as as other information technology resources, in collaboration with other institutions, for the benefit of the society.

In general, all those who promote research and use of intensive calculation, advanced communications technologies, as an instrument for sustainable socio economic development, devoting special attention to cooperative relations between public or private research centers and the productive sector.

3.8.1 Cloud Infrastructure

CESGA hosts a cloud infrastructure to provide an IaaS to the users. Currently, there are two infrastructures: one OpenNebula based for internal users and partners and other based in OpenStack affiliated to the EGI Federated Cloud (fedcloud). Several EGI Virtual Organizations (VOs) are supported and new ones can be added.

3.8.1.1 *Current status*

Regarding the OpenStack infrastructure, the Rocky version is used. Currently, a total of 336 cores and 756 GBytes of RAM are assigned to the federated cloud, but this amount of resources are in permanent expansion. A volume storage is available through the habitual OpenStack mechanism. This storage is hosted in a NetApp FAS9000 storage system.

3.8.1.2 *Further plans for integration*

The plans for integration include add support for the VOs related to the affiliate projects, add support for multiple OpenID identification providers and add more resources, both computation and storage.

3.8.2 Spark Big data

The Big Data infrastructure at CESGA provides several big Data related services, one of them is Spark. Spark is a fast and general engine for big data processing, with built-in modules for streaming, SQL, machine learning and graph processing.

3.8.2.1 *Current status*

Spark service at CESGA is currently offered to CESGA users and partners. The Big Data infrastructure at CESGA has a total of 38 nodes (4 masters and 34 slaves) connected with a 10GbE network and a total storage capacity of 816 TBytes.

3.8.2.2 *Further plans for integration*

CESGA wants to integrate its Spark service into the federated infrastructure. Details for the technological path to be chosen are being investigated. The goal is to open its resources to international research community. This is a challenge since there are few sites offering Spark in a federated framework.

3.8.3 CESGA HPC cluster

FinisTerrae-II is the name for the HPC cluster hosted at CESGA. FinisTerrae-II offers high performance computing services mainly to local and national researchers, but also to some international projects. It consists of a total of 320 computing nodes, 7.712 cores, 44.544 GB of memory and 750.000 GB of high performance storage (Lustre). All processing and computing nodes are interconnected through a Mellanox InfiniBand FDR low latency network. The peak computing capacity of the equipment is 328.272 Gflops and the sustained performance obtained in the Linpack test is 213.000 Gflops.

3.8.3.1 *Current status*

FinisTerrae-II has been providing hpc resources to the research community since 2016, and will be soon integrated in LIGO and IGFAE experiments. Currently it is not offers high performance computing services mainly to local and national researchers, but also to some international projects. It consists of a total of 320 computing nodes, 7.712 cores, 44.544 GB of memory and 750.000 GB of high performance storage (Lustre). All processing and computing nodes are interconnected through a Mellanox InfiniBand FDR low latency network. The peak computing capacity of the equipment is 328.272 Gflops and the sustained performance obtained in the Linpack test is 213.000 Gflops.

3.8.3.2 *Further plans for integration*

CESGA wants to integrate its HPC cluster in a bigger, federated infrastructure and tender its resources to international research projects. It is under study to address this as part of the European Open Science Cloud projects.

3.9 BIFI

The Institute for Biocomputing and Physics of Complex Systems (BIFI) of the University of Zaragoza develops research activities in different computing fields and provides services on different platforms: HPC, grid computing, cloud computing, GPUs, dedicated computers (FPGAs) and volunteer computing. The computing area of the institute plays a two-fold role, providing BIFI researchers and external organizations with computing resources and, at the same time, carrying out research into different fields of distributed/scientific computing. BIFI operates the Aragon node of Spanish Supercomputing Network (RES).

3.9.1 Cloud Facility

Colossus is a cloud-computing infrastructure based on OpenStack and has 1800 Intel Xeon E5-2670-v3 cores and a total of 20TB of RAM. Its internal and external connectivity is 10G Base T and its hypervisor in charge of virtualization is KVM. Furthermore, its storage solution is based on Ceph and has a raw capacity of 600TB.

The internal connectivity of the storage system is a high speed and low latency network (InfiniBand) and the external connectivity is 10Gbit/s. All these features allow a high data redundancy and availability. The

whole computing and storage infrastructure is highly redundant, load-balanced and monitored in order to guarantee a high availability to its users.

Colossus mainly offers a use of Infrastructure as a Service (IaaS), allowing users to manage their own machines and services. Colossus also allows a wide range of uses such as: highly CPU demanding scientific applications, corporate services, Big Data analysis and agile development and software integration, giving the possibility of creating small machines from 1 core and 1GB RAM to 24 cores and 240GB RAM.

3.9.1.1 *Current status*

Colossus is currently testing service integration with AAI dev and demo instances.

3.9.1.2 *Further plans for integration*

The next steps for service integration will be supporting AAI-demo instance, then going for AAI production instance, and following that, other services needed for the full FedCloud support will be integrated. Once that, colossus will support VOs by request. Furthermore we consider the integration with image management (AppDB), information service (BDII), monitoring, accounting, helpdesk, and registry (GOCDB).

3.10 CETA -CIEMAT

CETA-CIEMAT is a joint initiative with the regional government of Extremadura financed by PGE & FEDER. Our mission is to consolidate and disseminate eScience and IT, especially cloud, HPC and eInfrastructure to contribute to the effective expansion of eScience and facilitate usage of resources (new sites, new applications).

3.10.1 Cloud Facility

The OpenStack infrastructure currently runs the “Rocky” version. Currently, a total of 472 cores and 3904 GBytes of RAM are assigned to the federated cloud, but this amount of resources are in permanent expansion. A volume storage is available through the habitual OpenStack mechanism. This storage is hosted in a NetApp FAS3100 storage system with 97TB,

3.10.1.1 *Further plans for integration*

The plans for integration include add support for the VOs related to the affiliate projects, add support for multiple OpenID identification providers and add more resources, both computation and storage. We will also work on integration with the Core services described in section 2 (AAI, Monitoring, Accounting and Information System).

3.10.2 HPC Cluster

Turgalium is the name for the HPC cluster hosted at CETA-CIEMAT.

3.10.2.1 *Current status*

Turgalium offers high performance computing services mainly to local and national researchers, but also to some international projects. It consists of a total of 90 computing nodes, 1048 cores, 2752 GB of memory and 400 TB of high performance storage (Lustre). All processing and computing nodes are interconnected through a Mellanox InfiniBand FDR/QDR low latency network. Most nodes have a GPU coprocessor: 16 Tesla 1060, 16 Fermi 2050, 16 Fermi 2070 and 8 Kepler K80.

The HPC resources are accessible by a Slurm queue manager with a specific partition for ESOC-Synergy.

3.10.2.2 *Further plans for integration*

CETA-CIEMAT wants to integrate its HPC cluster with the federated infrastructure to provide its resources via additional interfaces to international user communities. It is currently being investigated which technologies to make use of.

4. Policy Gap Analysis

Accessing, using, and operating services for research, is inherently distributed. Users are expecting to access resources not located in their Home Organization. In this complex environment, the question of trust for both users and resource providers, or Infrastructures, becomes paramount.

To regulate and facilitate this trust, a set of policies is necessary. These policies outline the operation and operational measures undertaken by the Infrastructure to properly provide services. As such, the policies cover, among others, security measures, users' management and data protection.

This chapter intends to provide information about the necessary policies regulating access and usage of resources. It starts with the AARC Policy Development Kit (PDK) [AARC-Policy] but in addition also considers current best practices in research and education.

4.1 Introduction

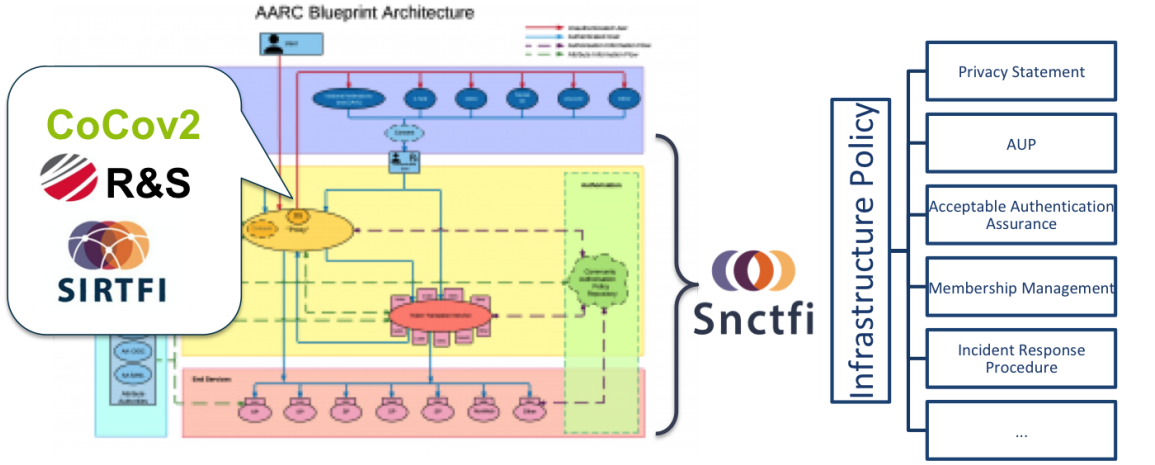
The AARC PDK considers the AARC Blueprint Architecture (BPA model), where typically user access is managed through an SP-IdP-Proxy component (or similar). The general components considered by the PDK (namely Data Protection, Security and Incident Response, and Membership Management) are relevant for managing authorization.

Policies are essential for setting expectations for participants in an Infrastructure, stretching from the Infrastructure management to the researchers themselves. Conversely, a violation of policy may be classified as a security incident and may warrant, and give grounds for, investigation to protect the Community. Policy decisions may or may not be enforced on a technical level; the Infrastructure themselves will be best placed to define the permitted usage of their resources through a combination of technology and documentation [AARC-PDK].

When incorporating external identities, Infrastructures are faced with questions, such as:

- Which policies do we need to legally receive attributes contained in federated identities?
- What is a reasonable expectation of the level of assurance of incoming identities?
- How can I ensure that all my users are covered by an incident response capability?
- What checks and measures should I put in place when managing the users of my community services, or members of virtual organisations?

The policies provided by the PDK aim to address these concerns by highlighting current best practices [AARC-PDK].



Armed with the recommended policy documents, and complying with best-practice community frameworks, an Infrastructure is able to support their users' activities in a federated environment (from AARC-PDK).

4.2 Policy Frameworks

The following frameworks are considered best practice for Research Communities enabling federated access. They enable trust and promote attribute release from the wider identity federation.

4.2.1 *Sirtfi Trust Framework*

Sirtfi demonstrates that an organisation complies with baseline expectations for operational security and incident response in the context of identity federations. All Service Providers and Identity Providers in Identity Federations are encouraged to support Sirtfi. To mitigate risk, the Infrastructure may choose to restrict its interactions to only those federated organisations who are able to comply with the framework. As well as the Infrastructure itself supporting Sirtfi, it is highly recommended that each connected service should support Sirtfi by providing their Security Contact and abiding by Sirtfi requirements. The Infrastructure should keep track of Security Contacts for connected services, ensure that participants are aware of the Security Incident Response Procedure and test preparedness for such a procedure regularly.

4.2.2 *Research and Scholarship Entity Category*

Research and Scholarship identifies federated services that are operated for the purpose of supporting research and scholarship activity. Identity Providers demonstrate their support for research and scholarship by releasing a defined set of attributes for a user, including name, email address and additional low-risk information that may be useful for their activities [R&S].

It is recommended that entities adopt and use this category since many Identity Providers will not release user attributes to services that do not publish the Research and Scholarship Entity Category.

4.2.3 *GÉANT Data Protection Code of Conduct*

The Data protection Code of Conduct (DPCoCo) describes an approach to meet the requirements of the EU Data Protection Directive and (version 2) with the General Data Protection Regulation (GDPR) in federated identity management. The Data protection Code of Conduct defines behavioral rules for Service Providers which want to receive user attributes from the Identity Providers managed by the Home Organisations.

It is highly recommended to use and abide by this framework, although there may be situations where the DPCoCo is not applicable due to having a more restrictive policy in place or legal impediments to adoption. DPCoCo will provide a scalable and harmonised approach when processing users personal data. Whenever users are accessing services, facilitated by the release of users' information and its subsequent processing by the service providers, which is the most common scenario in federated environment, there needs to be a legal basis for such processing. DPCoCo aims to provide such a basis.

Currently, DPCoCo is still in the process of consultation. However, once the final version of the document is submitted for consideration by Data Protection Authorities (national Data Protection Agency or EU one), we strongly recommend for all IdPs and all SPs to adopt it and adhere to it. More information can be found here (home page of the overall effort) and here (new DPCoCo version). Explanatory memorandum offers more explanation about the reasons for CoCo and its approach.

4.3 AARC Policy Development Kit

This chapter covers the policies contained in the policy kit, their meaning, purpose, and possible application.

The Policy Kit builds on the Snctfi framework [SNCTFI]. The top level Infrastrastructure Policy serves to bind the entire policy set and stipulates requirements on each of the participants; Management, Infrastructure Security Contact, User Community Management, Service Management (including the Proxy Operator) and the User. The top policy identifies additional policy documents; in this case the five that are mandatory for Snctfi compliance. The Infrastructure may wish to define additional policies, such as Service Eligibility, Disaster Recovery, or Data Management; these policies should be linked into the Infrastructure Policy to ensure a coherent Policy set.

		Manage ment	Infrastructure Security Contact	User Community Management	Service Management	User
Top Level	Infrastructure Policy	Defines & Abides by	Abides by	Abides by	Abides by	
Data Protection	Privacy Statement	Defines			Defines	Views
	Policy on the Processing of Personal Data	Defines	Abides by	Abides by	Abides by	
Membership	Community	Defines		Abides by		

Management	Membership Management Policy					
	Acceptable Use Policy	Defines		Defines		Abides by
	Acceptable Authen-tication Assurance	Defines		Abides by	Abides by	
Operational Security	Incident Response Procedure	Defines	Abides by		Abides by	
	Service Operations Security Policy	Defines			Abodes by	

Top Level Policy regulates the behavior and activities of participants in the Infrastructure, and binds all other policies in a coherent whole. It explains the relevant terms, and instructs certain actions to be taken. The Infrastructure must have a Security Officer. All services must have a designated Security Contact. The communities must designate a Security Contact, and must ensure that all Community users will accept and abide by the relevant policies (which are all policies). This can be achieved, for example, by showing an Acceptable Use Policy (AUP) that contains links to all Infrastructure Policies. Naturally, this can technically be done by the Infrastructure's services.

Membership Management Policy is a set of rules for the Community on how User membership should be managed. The Community must define an AUP. The template is provided. The Community must properly manage their users' membership life cycle, and must record all actions conducted on it. All the outlined actions must be followed (i.e. rules for Registration, Assignment of Attributes, Renewal, Suspension, Termination). The Community must take actions to ensure proper data protection and auditability.

Acceptable Authentication Assurance Policy outlines the acceptable authentication assurance for the community, but also for the Infrastructure. The standard way of conveying this information is to use the REFEDS Assurance Framework (RAF) [RAF]. The Community must define their own Assurance procedures, especially in relation to Identity Vetting. This may depend on the acceptable assurance levels demanded by services, e.g. services may request RAF Assurance Profile Cappuccino, and Community Manager must ensure that it is followed.

Acceptable Use Policy defines conditions of usage of Infrastructure resources, but may additionally define rules for the Community itself. At the very least, Community must input their name and purpose. The Community may reuse the Infrastructure policy, if that is enough for them.

Policy on the Processing of Personal Data outlines that proper measures must be taken to protect the personal data of users when using Infrastructure services, but it also instructs the Community to do the same. The Community must accept this policy, and must ensure that, if Community has services integrated with the Infrastructure, must follow these rules.

Privacy Policy Template is a template for all the services to use and follow.

Incident Response Procedure is a set of rules to follow in case of a security incident. All Services must follow and abide by this procedure.

These policies and their templates can be found at the AARC website [AARC-Policy].

Additionally, there is a Moodle course that serves as an introduction for the PDK, and explains the purpose and usage of policies. The course allows one to organise and systematise the policy writing and implementation with the Infrastructure in order to properly manage users and properly provide services [PDK-MOODLE]. Everyone that needs to understand or create policies in federated research context is strongly encouraged to take the course. The course is also available as a YouTube playlist [PDK-Playlist].

4.4 Current overview of the support for policies

A set of questions was prepared as part of this work package effort and sent to EOSC-Synergy sites (resource providers) in order to identify their level of preparedness and interoperability with the general EOSC efforts in terms of trust and security. This question is asked in terms of their operational procedures (or policies), not in terms of particular resources that are provided. These questions relied on the policy and policy frameworks information provided above, and it was structured as to follow the main structure of the PDK.

The following organisations participated in the survey:

- Institute of Informatics SAS
- Supercomputing Center of Galicia (CESGA)
- Karlsruhe Institute of Technology
- AGH CYFRONET
- BIFI - UNIZAR
- LIP Lisbon
- CSIC
- LIP
- CETA-CIEMAT
- CESNET
- CSIC
- Data Archiving and Networked Services

In total, 12 organizations participated (with 13 participants in total), with the services detailed in section 3. The questions were oriented around 3 major areas, i.e. Operational Security, Data Protection, and support and awareness of Policy Frameworks (Appendix C). Detailed answers to all the questions can be found in Appendix D.

The highlighted results are the following:

1. More than half of the participants (7 out of 13, or 54%) have not heard about the AARC Policy Development Kit. This may indicate that greater effort is needed to disseminate current best efforts and practices in order to support interoperability.
2. More than 60% have heard of the Sirtfi framework (8 out of 13), however only 38% (5 out of 13) indicated that they do have a formal Incident Response Procedure.

3. Even with the lack of formal Incident Response Procedure, almost 70% of participants (9 out of 13) have a security contact for their services, and more than half (7 out of 13) have a security policy for the service they are operating.
4. Most, but not all (10 out of 13) people are aware of the GDPR, and most (8 out of 13) have a data protection policy for their institution. However, only 2 out of 13 people confirmed the existence of privacy notice for the service they are operating.
5. Most participants (9 out of 13) have at least some kind of support (in terms of manpower) for integrating their services in EOSC, and running the service in production.

4.5 Summary

Accessing, using, and operating services for research in today's world, as a rule, is inherently distributed. In order to regulate and facilitate this trust, a set of policies is necessary [AARC-PDK]. Various policies and best operational procedures are already available for wider adoption in the Research and Education activities, especially regarding operating services and managing and regulating the access to those services. This document outlines the relevant frameworks, like [AARC-Policy], and provides instructions on how to use and apply the policies [PDK-MOODLE, AARC-Policy, AARC-Guidelines]. Current efforts within EOSC and other supporting projects will continue to increase the adoption and harmonization of the policies, a necessary undertaking to provide interoperability within EOSC participants.

5. Gender Issues

The project management is completely engaged with promoting a working atmosphere free of any type of discrimination for gender reasons. In this sense no issues were identified during the proposal preparation time nor during the initial stage of the project.

The Project Management will make sure that the selection of profiles to participate in the different bodies of the project will take place by evaluation exclusively the relative merits of the candidates from a curricular point of view.

6. Conclusions

The deliverable provides a detailed overview of the progress and achievements during the first reporting period of the project. It divides the roadmap into integration on three different levels and underlines the current status of integration work for EOSC-Synergy infrastructure services with EOSC-hub federated services.

The document first described the fundamental EOSC-hub services and identified core services by their functions in order to integrate and provide initial procedures on how to integrate with each one. The analysis reveals a few technical solutions that are relevant to most of the identified services.

A second important aspect of the deliverable was to provide national capacity integration status. It provided an overview of the infrastructure services in terms of their architectural type, covering a description of the service. As a result of this analysis, the levels of integration with the EOSC core

services were identified and documented. The document contains information about the current status and further plans for the integrations. The status can be summarised as the first level being largely already accomplished, work on the second level is being started, and the procedures for integration on the third level are closely monitored.

Finally, the deliverable goes through the Policy Gap Analysis, where various policies and best operational procedures are identified, and instructions on how to use and apply these policies are provided.

The deliverable contains an annex with an overview of the integration status and plans for each one of the infrastructure services.

7. References

[AAI-Integration-Procedure] Integration procedure for EOSC AAI
https://docs.google.com/spreadsheets/d/1ZCGXjt6x3eZ3V4TnmyysfBZoRMl4hBK2urpZhoYpNyQ

[AAI101] Authentication-and-Authorisation-101-tutorial
https://aarc-project.eu/wp-content/uploads/2019/03/Authentication-and-Authorisation-101-tutorial.pdf

[AAI-Advanced] Introduction to advanced AAI components
https://aarc-project.eu/wp-content/uploads/2019/03/Introduction-to-advanced-AAI-components.pdf

[AARC-Guidelines] AARC Policy Guidelines
https://aarc-community.org/guidelines/#policy

[AARC-G021] Exchange of specific assurance information between Infrastructures
https://aarc-project.eu/guidelines/aarc-g021/

[AARC-G048] Guidelines for Secure Operation of Attribute Authorities and other issuers of access statements
https://aarc-project.eu/guidelines/aarc-g048/

[AARC-Policy] AARC Webpage on Policies
https://aarc-community.org/policies/policy-development-kit

[AARC-PDK] AARC Policy Development KIT (PDK)
https://docs.google.com/document/d/176vzNaoK6KvKTMp8Glk2n1NaM6bxiS1QqH8M3_mu7NI/edit#

[EGI-FedCloud] EGI Federated Cloud Homepage and architecture description
https://www.egi.eu/federation/egi-federated-cloud

[EGI-FedCloud-Architecture] Architecture of the EGI Federated Cloud
https://wiki.egi.eu/wiki/Federated_Cloud_Architecture

[EGI-Proc-09] Resource Centre Registration and Certification
https://wiki.egi.eu/wiki/PROC09_Resource_Centre_Registration_and_Certification

[EGI-Proc-14] Process of enabling a Virtual Organisation (VO) on the Infrastructure
https://wiki.egi.eu/wiki/PROC14_VO_Registration

[EOSC core and common] EOSC-hub federated services list
https://docs.google.com/spreadsheets/d/1_uFvUPQMxr619L9H2cmHB7CcbcG91F_xgSGr7JJYhiQ

[EOSC Onboarding] EOSC-hub onboarding procedures
https://eosc-portal.eu/for-providers

[GDPR] General Data Protection Regulation
https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679

[KANTARA] Kantara Identity Assurance Framework is a set of controlling documentation
https://kantarainitiative.org/confluence/display/LC/Identity+Assurance+Framework

[PDK-MOODLE] AARC PDK Moodle course
https://e-academy.geant.org/moodle/course/view.php?id=16

[PDK-Playlist] AARC Project YouTube list
https://www.youtube.com/playlist?list=PLELuOn8jN3IIbp0W-WxO6712JKGz7qK0N

[R&S] REFEDS Research and Scholarship
https://refeds.org/category/research-and-scholarship

[RAF] REFEDS Assurance Framework
https://wiki.refeds.org/display/ASS/Assurance+Home

[SNCTFI] Scalable Negotiator for a Community Trust Framework in Federated Infrastructures
https://aarc-community.org/policies/snctfi/

Appendix A - Acronyms

Acronym	Description
AAI	Authentication and Authorisation Infrastructure
AARC	Authentication and Authorisation for Research and Collaboration, EU Project H2020 Grants 653965 and 730941)
AARC PDK	AARC Policy Development Kit
AGH	Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie
API	Application Program Interface
AUP	Acceptable Use Policy
BDII	Berkeley Database Information Index
BIFI	Biocomputación y Física de Sistemas Complejos
CESGA	Centro de Supercomputación de Galicia
CESNET	Zajmove Sdruzeni Pravnickych osob
CETA-CIEMAT	El Centro Extremeño de Tecnologías Avanzadas
CMDB	Configuration Management Database
CSIC	Consejo Superior de Investigaciones Científicas
DPCoCo	Data protection Code of Conduct
DPMT	Data Project Management Tool
EGI	European Grid Initiative
EOSC	European Open Science Cloud
FEUDAL	Federated User Deployment portal

FPGA	A field-programmable gate array
GDPR	General Data Protection Regulation
GOCDB	Grid Configuration Database
GPU	Graphics Processing Unit
HPC	High Performance Computing
HTC	High Throughput Computing
IaaS	Infrastructure as a Service
IFCA	Instituto de Física de Cantabria
IISAS	Independent, Intelligent, Solutions and Services
IM	Infrastructure Manager
INCD	Infraestrutura Nacional de Computação Distribuída
KIT	Karlsruher Institut für Technologie
LSDF	Large Scale Data Facility
OIDC	OpenID Connect
PaaS	Platform as a Service
PSNC	Poznan Supercomputing and Networking Center
RAM	Random-access memory
REBUS	WLCG REsource, Balance & USage
SaaS	Software as a Service
SNCTFI	Scalable Negotiator for a Community Trust Framework in Federated Infrastructures
SGE	Son of Grid Engine
Sirtfi	Security Incident Response Trust Framework for Federated Identity
SPMT	Service Portfolio Management Tool
SQL	Structured Query Language
SVB	Service Validation Board
UID	User Identifier
VOMS	Virtual Organization Membership Service
VOs	Virtual Organizations

WebDAV	Web Distributed Authoring and Versioning

Appendix B - Service Integration Status

Service	Service Integration				
	AAI	Monitoring	Accounting	Info Provider	Service Registry
KIT (Storage facility)	Planned	Planned	Planned	Planned	Considered
KIT (WaTTS)	In progress	Planned	Does not apply	Does not apply	Planned
KIT (FEUDAL)	Done	Planned	Does not apply	Does not apply	Planned
CSIC (Cloud Infrastructure)	Done	Done	Done	Done	Done
LIP (INCD NCG CLOUD)	In progress	Undergoing	Undergoing,	Undergoing	Done
LIP / INCD NCG FARM	GRID with VO/VOMS	Done	Done	Done	Done
LIP / INCD RDA CLOUD	Planed	Planed	Planed	Planed	Planed
LIP / INCD RDA FARM	Planed	Planed	Planed	Planed	Planed
CYFRONET	Done	Done	Does not apply	Does not apply	Does not apply
PSNC Cloud	In Progress	In progress	In progress	In progress	In progress
PSNC HTC	Done	Done	Done	Done	Done
IISAS	Done	Done	Done	Done	Done
CESNET (Cloud Infrastructure)	Done	Done	Done	Done	Done
CESNET	In progress	Done	Done	Done	In progress
CESGA	Planed	Planed	Planed	Planed	Planed
BIFI	In progress	Planed	Planed	Planed	
CETA -CIEMAT	In progress	In progress	In progress	In progress	

Table 1. Summary of the service integration status

Appendix C - Policy Questionnaire

List of questions sent in the Policy Gap Analysis:

Operational Security

1. Are you aware of Sirtfi?
2. Do you have an Incident Response Procedure?

3. If yes, is it Sirtfi compliant?
    a. Do you have a CSIRT?
    b. Do you have a designated security person contact?
    c. If yes, is it publicly available?
4. Do you have a general Security Policy (valid site-wide)?
5. Do you have a Security Policy per Service?
6. If yes, how compatible is it with PDK [AARC-Policy]?
7. Are you running Services that require more secure operation (e.g. Attribute Authority, DB storing private data, e.g. LDAP, credential issuance services, e.g. Certificate Authorities)?
8. If so, do you have a policy regulating how secure they are operated?
9. If yes, how compatible is it with the AARC-G048 [AARC-G048] guideline?

Membership Management

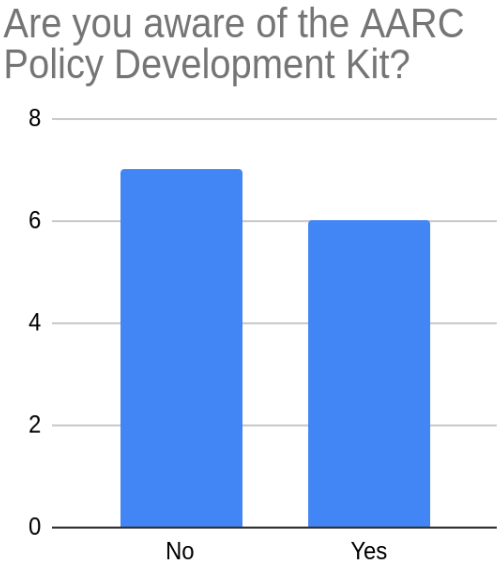
1. Are you running services for a Community to manage their membership (e.g. Attribute Authority or IdP-SP-Proxy, described in [AAI101] and [AAI-Advanced])?
2. If yes, do you have a Membership Management Policy that regulates how the community must manage their users (e.g. onboarding, suspension, removal of users)?
3. If yes, is it compatible with the PDK [AARC-Policy]?
4. Do you have an Acceptable Use Policy (AUP)?
5. If so, is it compatible with the PDK?
6. Do you have a policy regulating an Acceptable Authentication Assurance, regulating, e.g., identity vetting assurance, attribute quality and freshness, and similar?
7. If yes, is it compatible with the REFEDS Assurance Framework, RAF [RAF]?

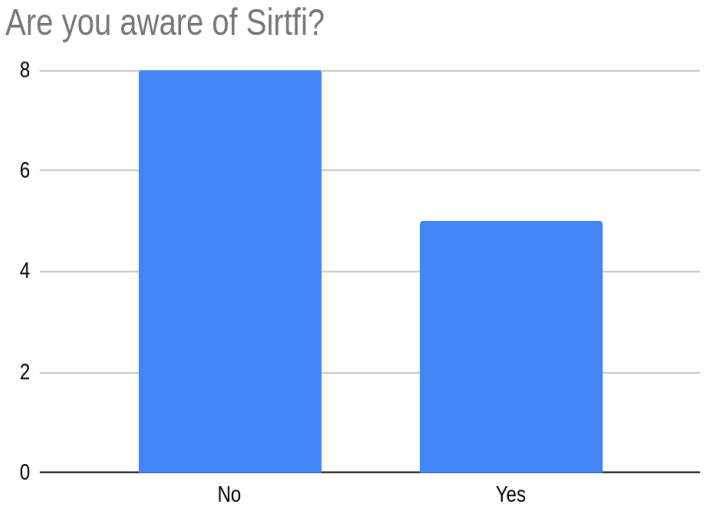
Data Protection

1. Are you aware of the General Data Protection Regulation, GDPR [GDPR]?
2. If yes, do you have a general Data Protection Policy (site-wide)?
3. If yes, is it compatible with the PDK?
4. Do you have a service-based Privacy Notice?
5. If yes, is it compatible with the PDK?
6. Have you documented what kind of personal data is processed?
7. Have you conducted a Risk Assessment in relation to the processing of personal data?
8. If yes, have you documented it?

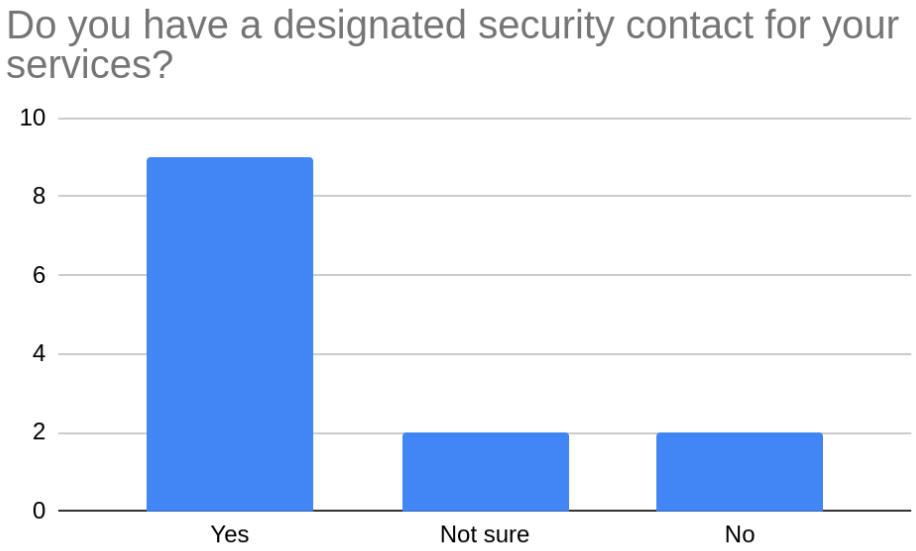
Appendix D - Questionnaire Results

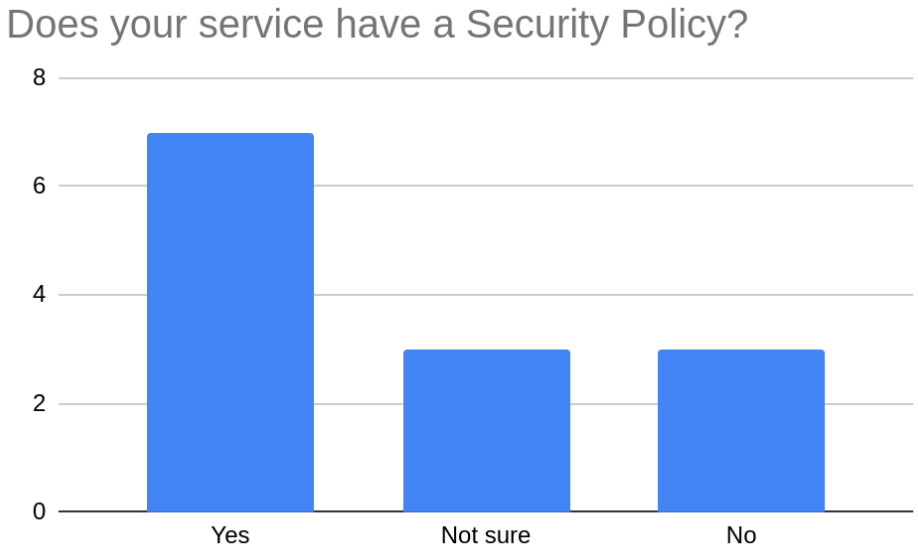
Answers to all questions in the chapter 4, Policy Gap Analysis.

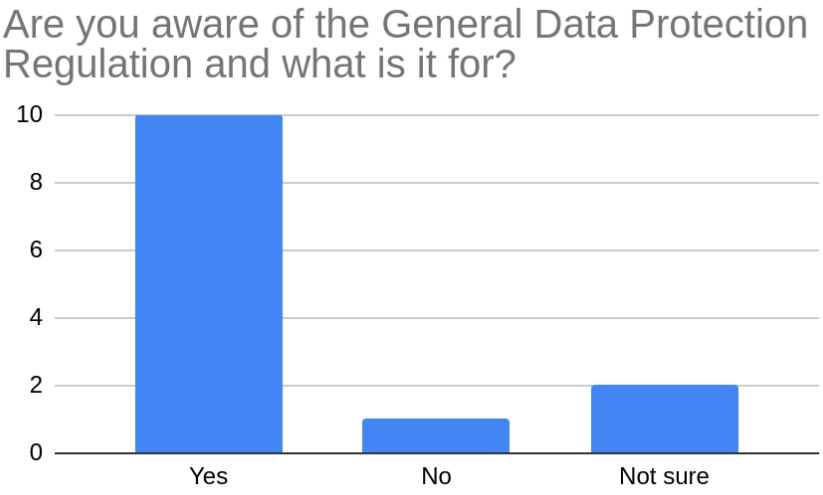


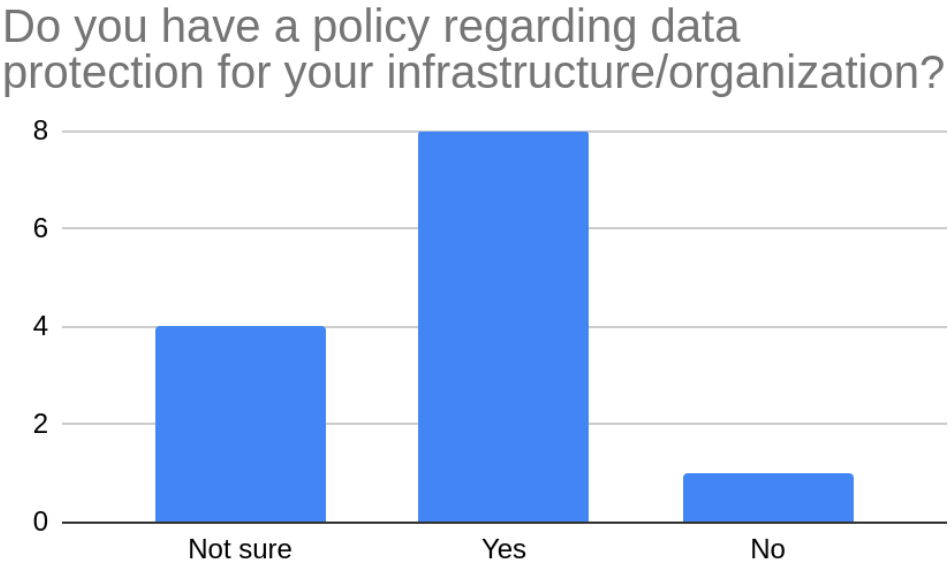


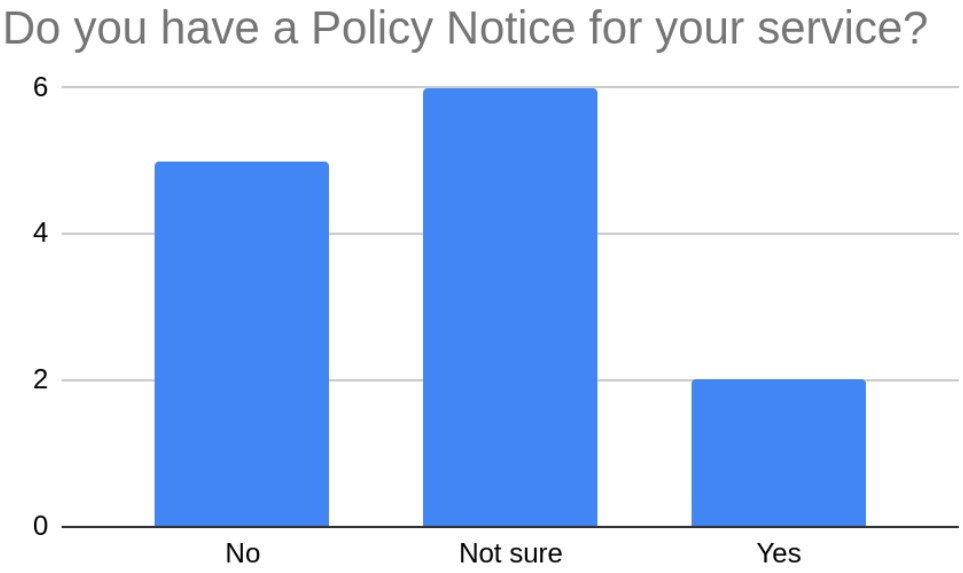

Does your site have an Incident Response Procedure?
5
4
3
2
1
0
Yes
Not sure
Doesn't apply
No

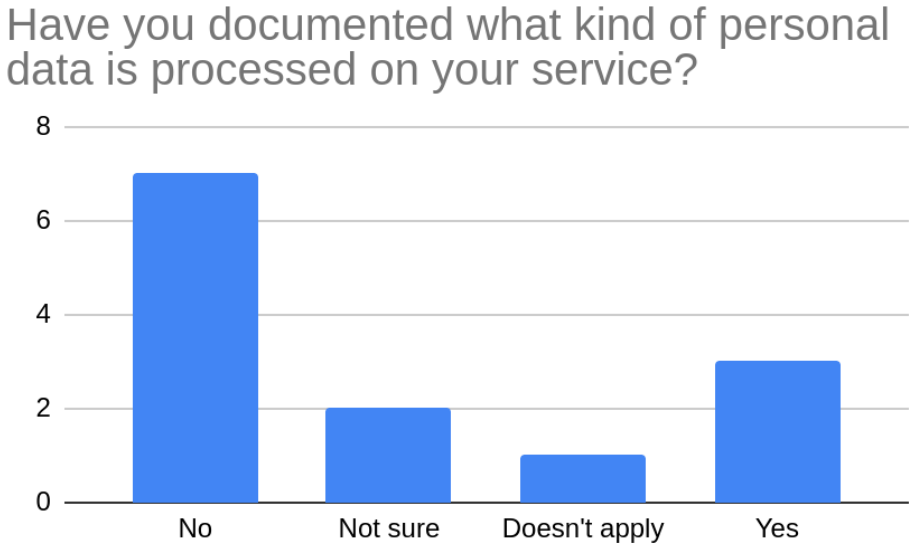

Do you have a designated security contact for your services?
10
8
6
4
2
0
Yes
Not sure
No

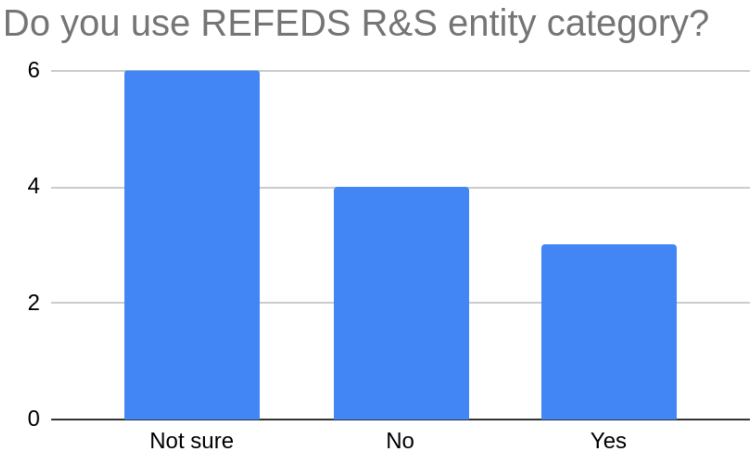

Does your service have a Security Policy?
8
6
4
2
0
Yes
Not sure
No


Are you aware of the General Data Protection Regulation and what is it for?
10
8
6
4
2
0
Yes
No
Not sure

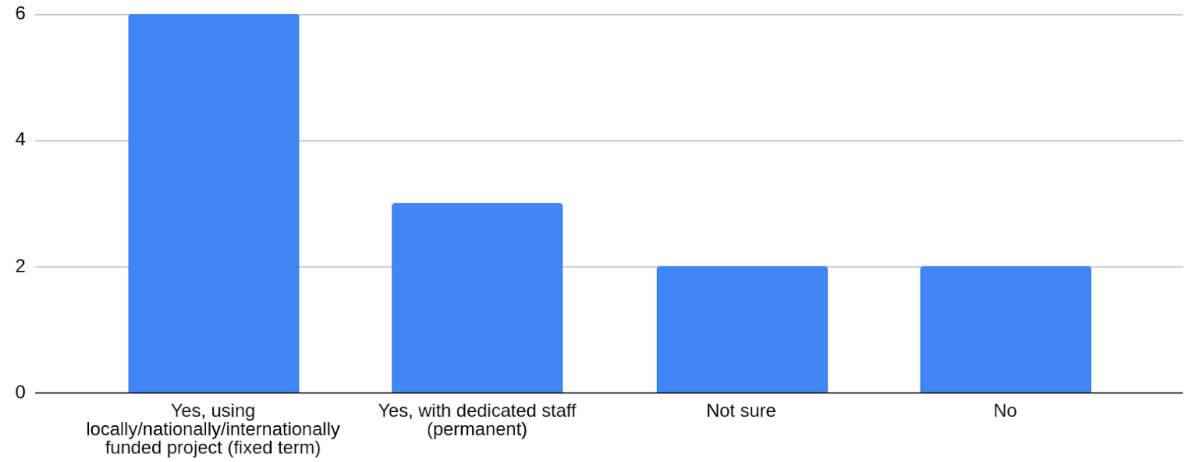

Do you have a policy regarding data protection for your infrastructure/organization?
8
6
4
2
0
Not sure
Yes
No


Do you have a Policy Notice for your service?
6
4
2
0
No
Not sure
Yes


Have you documented what kind of personal data is processed on your service?
8
6
4
2
0
No
Not sure
Doesn't apply
Yes


Do you use REFEDS R&S entity category?
6
4
2
0
Not sure
No
Yes

Does your organisation provide a support for integrating the services in EOSC, together with the effort of running the service in production?



Do you use any assurance framework?